THE SEISMOLOGICAL SOCIETY OF AMERICA
400 Evelyn Ave., Suite 201
Albany, CA 94706-1375
(510) 525-5474; FAX (510) 525-7204
www.seismosoc.org

Helmstetter, A., & Werner, M. J. (2012). Adaptive Spatiotemporal Smoothing of Seismicity for Long-Term Earthquake Forecasts in California. Bulletin of the Seismological Society of America, 102(6), 2518-2529. 10.1785/0120120062

Bulletin of the Seismological Society of America, Vol. 102, No. 6, pp. 2518–2529, December 2012, doi: 10.1785/0120120062

Adaptive Spatiotemporal Smoothing of Seismicity for Long-Term Earthquake Forecasts in California

by Agnès Helmstetter and Maximilian J. Werner

Abstract We present new methods for time-independent earthquake forecasting that employ space–time kernels to smooth seismicity. The major advantage of the methods is that they do not require prior declustering of the catalog, circumventing the relatively subjective choice of a declustering algorithm. Past earthquakes are smoothed in space and time using adaptive Gaussian kernels. The bandwidths in space and time associated with each event are a decreasing function of the seismicity rate at the time and location of each earthquake. This yields a better resolution in space–time volumes of intense seismicity and a smoother density in volumes of sparse seismicity. The long-term rate in each spatial cell is then defined as the median value of the temporal history of the smoothed seismicity rate in this cell. To calibrate the model, the earthquake catalog is divided into two parts: the early part (the learning catalog) is used to estimate the model, and the latter one (the target catalog) is used to compute the likelihood of the model's forecast. We optimize the model's parameters by maximizing the likelihood of the target catalog. To estimate the kernel bandwidths in space and time, we compared two approaches: a coupled near-neighbor method and an iterative method based on a pilot density. We applied these methods to Californian seismicity and compared the resulting forecasts with our previous method based on spatially smoothing a declustered catalog (Werner *et al.*, 2011). All models use small $M \geq 2$ earthquakes to forecast the rate of larger earthquakes and use the same learning catalog. Our new preferred model slightly outperforms our previous forecast, providing a probability gain per earthquake of about 5 relative to a spatially uniform forecast.

Introduction

Many long-term earthquake forecasts are based on smoothed seismicity (e.g., Kagan and Jackson, 1994; Frankel *et al.*, 1997; Rong and Jackson, 2002; Stock and Smith, 2002a,b; Helmstetter *et al.*, 2006, 2007; Werner, Helmstetter, *et al.*, 2010; Zechar and Jordan, 2010; Werner *et al.*, 2011). These studies suggest that smoothing the locations of past earthquakes provides a skillful estimate of the spatial distribution of future seismicity. However, a recurring problem with this approach is the strong clustering of seismicity in time. Seismicity rates can increase by several orders of magnitude after large earthquakes, which would lead to strong peaks in a spatial forecast if the temporal clustering were not removed.

For instance, Werner *et al.* (2011, their fig. 3) found that declustering a Californian catalog prior to smoothing epicenters resulted in larger probability gains than smoothing all quake locations. It seems reasonable to expect that their findings extend to many other regions and catalogs, especially for small to moderate magnitude thresholds. This provides quantitative support for the decision by almost all authors to decluster prior to smoothing to obtain better long-term forecasts. There are, however, some exceptions. For instance, Zechar and Jordan (2010) smoothed an entire Italian catalog to forecast seismicity, and their forecasts often performed well in retrospective tests (Werner, Zechar, *et al.*, 2010). But Zechar and Jordan (2010) did not test whether smoothing a declustered catalog would have produced better forecasts using their method.

Despite the advantage declustering seems to offer for forecasts, however, declustering an earthquake catalog is, to some extent, subjective and, moreover, a retrospective enterprise of some controversy (e.g., Van Stiphout *et al.*, 2011). Several algorithms have been proposed; the more frequently used are the methods by Gardner and Knopoff (1974) and by Reasenberg (1985). Each method has several adjustable parameters, which may significantly affect the results (Van Stiphout *et al.*, 2011), but are generally chosen in a somewhat arbitrary manner. Another class of methods does not classify earthquakes as either belonging to the background or to the triggered events, but assigns a probability of independence to each earthquake (Kagan, 1991; Zhuang *et al.*, 2002, 2004; Marsan and Lengliné, 2008). These probabilistic

methods have the advantage that parameters are optimally chosen (e.g., using a maximum likelihood approach) rather than arbitrarily fixed. Nonetheless, both deterministic and stochastic declustering methods make several strong assumptions about seismicity.

Declustering methods generally assume that the seismicity rate is the sum of a constant loading rate and triggered earthquakes. Therefore, these methods cannot account for quiescence, that is, a decrease of seismicity below the background rate often claimed to occur in Coulomb stress shadows following large mainshocks (e.g., Toda *et al.*, 2012) or variable loading rates (e.g., due to fluid flow or aseismic slip). Moreover, the choice of declustering method might influence the performances of the long-term forecast.

Few alternatives to declustering algorithms have been proposed. Hainzl *et al.* (2006) suggested to estimate the fraction of background events from the interevent time distribution. This method can be used to map the background rate, however, with a limited resolution in space. Stock and Smith (2002a,b) used spatial kernels to smooth seismicity in separate time intervals of one year. They then estimated the rate of independent events from the statistics of the density of seismicity in each time interval. Specifically, they computed the time series $R_{t,i}$ of the rate in each cell i and for each year t, as well as the rate R_i using the full time window. To correct for temporal clustering, the rate R_i is divided by $(1 + V_i)^2$, where V_i is the index of dispersion (variance-to-mean ratio) of the time series $R_{t,i}$. This lowers the predicted rate in cells where many aftershocks have occurred and $V_i \gg 1$. Our method is similar to that of Stock and Smith (2002a,b) in that we estimate the density of seismicity using adaptive kernels and then estimate the stationary rate from the temporal variability of the density. The main differences are (1) we use adaptive kernels in space and time, while Stock and Smith (2002a,b) used adaptive kernels in space only in distinct time intervals, and (2) we estimate the long-term rate from the median value of the seismicity rate in each spatial cell rather than from the index of dispersion of the number of events.

In this study, we are only interested in time-independent earthquake forecasts, although the methods presented here can also be used for time-dependent, short-term forecasts, which we will present in a separate article. Below, we first present the adaptive kernels used to estimate the density of earthquakes in space and time and describe two methods for estimating the spatial and temporal bandwidths: the coupled near-neighbor method and an iterative method based on a pilot estimate. We then discuss how these methods can be used to provide earthquake forecasts. Finally, we compare the forecasts with the long-term forecast by Werner *et al.* (2011), which is based on adaptive smoothing of declustered seismicity. We have submitted this forecast to the Collaboratory for the Study of Earthquake Predictability (CSEP), whose mission is to provide the infrastructure for regionally and globally testing prospective earthquake forecasts (e.g., Jordan, 2006; Zechar *et al.*, 2010).

Spatiotemporal Estimation of the Seismicity Rate using Adaptive Kernels

Nonparametric kernel estimation is often used in seismology to map the spatial distribution of seismicity and provide long-term forecasts. Most studies use kernels in the spatial domain only, using either fixed kernels with a constant bandwidth (e.g., Kagan and Jackson, 1994; Frankel *et al.* 1997; Zechar and Jordan, 2010) or with adaptive bandwidths (e.g., Stock and Smith, 2002a,b; Helmstetter *et al.*, 2006, 2007; Werner *et al.*, 2011). Adaptive kernels allow for a higher resolution in areas of dense seismicity, while avoiding isolated spikes in zones of sparse seismicity. Few studies have used kernels in both the spatial and temporal domain (Choi and Hall, 1999; Adelfio and Chiodi, 2010). The intensity at location $\mathbf{r}$ and time t can be expressed as a function of the times and locations of earthquakes in a catalog by

$$R(\mathbf{r}, t) = \sum_{i=1}^{N} \frac{1}{h_i d_i^2} K_1\left(\frac{t - t_i}{h_i}\right) K_2\left(\frac{|\mathbf{r} - \mathbf{r_i}|}{d_i}\right), \qquad (1)$$

where h_i and d_i are the bandwidths associated with event i in the temporal and spatial domain, respectively, and $|\mathbf{r} - \mathbf{r_i}|$ is the distance to the epicenter of event i. Different functions (e.g., power laws) could be used for the temporal kernel K_1 and the spatial kernel K_2. Here, we choose Gaussian kernels for both K_1 and K_2 for simplicity. Adaptive kernels assume that h_i and d_i vary as a function of the density at the location and time of each earthquake. Because of the limited accuracy of earthquake locations, we impose that the spatial bandwidth d_i cannot be smaller than 0.5 km. Because we are interested in earthquake forecasting (rather than a retrospective density estimation), we use a modified expression of equation (1), so that the seismicity rate at time t only depends on past earthquakes with $t_i < t$

$$R(\mathbf{r}, t) = R_{\min} + \sum_{t_i < t} \frac{2}{h_i d_i^2} K_1\left(\frac{t - t_i}{h_i}\right) K_2\left(\frac{|\mathbf{r} - \mathbf{r_i}|}{d_i}\right). \qquad (2)$$

The factor two in equation (2) accounts for the weight of future events in the sum of equation (1). The constant parameter $R_{\min}$ assures that R is always positive, thereby accounting for surprises, that is, earthquakes that occur in areas without prior seismic activity. Two methods are available to estimate the bandwidths d_i and h_i associated with each earthquake, the coupled near-neighbor method and the iterative method based on a pilot estimate.

Coupled Near-Neighbor Method

In seismology, the kernel bandwidths in space and/or in time are usually estimated from the distance between each earthquake and its nearby neighbors (e.g., Choi and Hall, 1999; Zhuang *et al.*, 2002; Helmstetter *et al.*, 2006, 2007; Adelfio and Chiodi, 2010; Werner *et al.*, 2011). For each earthquake, the spatial bandwidth d_i is estimated as

the distance between earthquake i and its kth nearest neighbor, where k is a fixed integer (to be optimized). In order to measure both the spatial and temporal bandwidths d_i and h_i, the simplest method is to estimate d_i and h_i independently (Adelfio and Chiodi, 2010), by defining h_i as the minimum of $(t_{i+k} - t_i)$ and $(t_i - t_{i-k})$. When applied to earthquake forecasting, d_i and h_i should only depend on past events $j < i$, so that $h_i = t_i - t_{i-k}$. Estimating h_i and d_i independently, as done by Adelfio and Chiodi (2010), suffers from a major drawback: events that occurred at the same location as a major aftershock sequence, but long before or after the mainshock, will have a very small spatial footprint d_i because many events occur nearby, even if the density of seismicity at the time of the event is low.

A better, coupled method was proposed by Choi and Hall (1999), which we modify for asymmetric temporal kernels. For each earthquake i, the bandwidths d_i and h_i are chosen to minimize the function $h_i + ad_i$ under the constraint that there are at least k events at a distance smaller than d_i and, simultaneously, at a time between $t_i - h_i$ and t_i. This method is illustrated in Figure 1. The parameter k controls the overall level of smoothing, while a controls the relative importance of smoothing in space versus time. A larger value of a gives a forecast with a higher resolution in space that is simultaneously smoother in the time domain.

Iterative Method Based on a Pilot Estimate

Another approach to estimate the optimal adaptive bandwidths in equation (2) consists in evaluating iteratively the bandwidth as a function of a pilot-estimate $\hat{R}$ of the spatiotemporal seismicity density. Abramson (1982) showed desirable statistical properties of the resulting density estimate (e.g., vanishing bias) if the bandwidths scale according to $\sim 1/\sqrt{\hat{R}(x, y, t)}$. This is true for both univariate and multivariate kernels. Nearest-neighbor bandwidths, in contrast, scale as $\sim 1/\hat{R}(x, y, t)$ (in 1D), that is, they are more strongly dependent on the local density. Abramson (1982) also noted that, in 2D, the two dependencies agree. Nonetheless, here we are concerned with the practical issues of estimation and the predictive skill of the resulting density; therefore, we compare the two approaches. The pilot estimate can be obtained by using equation (2) with uniform bandwidths h and d (Silverman, 1986; Stock and Smith, 2002a,b). We find that the earthquake forecasts based on a pilot estimate improve when iterating this process several times. Starting from uniform values ($h_i = h_0$, $d_i = d_0$), $\hat{R}(t_i, r_i)$ at the time and location of each earthquake is estimated using (2), and the modified bandwidths are estimated by

$$h_i = h_0 \sqrt{\frac{< \hat{R}(t_i, r_i) >_{\text{geo}}}{\hat{R}(t_i, r_i)}}, \tag{3}$$

and

$$d_i = d_0 \sqrt{\frac{< \hat{R}(t_i, r_i) >_{\text{geo}}}{\hat{R}(t_i, r_i)}}, \tag{4}$$

where the symbol $<>_{\text{geo}}$ denotes the geometrical average over all events $i = 1, \ldots, N$ in the learning catalog. After each iteration, the log-likelihood of the estimated (and normalized) seismicity density L_R is estimated by

$$L_R = \sum_{i=1}^{N} \log[\hat{R}(t_i, r_i)]. \tag{5}$$

If the likelihood after one step is larger than after the previous step, the bandwidths are reestimated using equations (3) and (4), the seismicity rate $\hat{R}(t_i, r_i)$ is updated using equation (2), and the likelihood is estimated using equation (5) until L_R becomes smaller than its previous value. Typically, we find that this process stops after two to five iterations.

Long-Term Forecast

The spatiotemporal seismicity density estimated by adaptive kernels in equation (2) is then used to estimate the stationary long-term rate at each node of a spatial grid and to produce long-term earthquake forecasts in California.

Data: The ANSS Earthquake Catalog

We use the Advanced National Seismic System (ANSS) catalog in the period from 1 February 1981 until 22 July 2011 with magnitude $M \geq 2$ and depth smaller than 100 km. This is a composite catalog that gathers data from different seismic networks in the United States. Events likely to be associated with underground nuclear explosions at the Nevada Test Site have been removed following Werner *et al.* (2011). There are also problematic events that appear twice (or more often) in the catalog; these are mostly small events. For instance, there are eight nearly identical events listed in the ANSS catalog within 0.2 s on 20 August 2003, with about the same location and magnitude, suggesting a catalog artifact. We did not attempt to remove all the duplicate events; we only removed 68 events that had exactly the same time, location, and magnitude as the preceding event. This leaves 188,319 events with latitudes ranging between 30.5° N and 44° N and longitudes between 126.4° W and 112.1° W. The catalog is reasonably complete above M 2 in the central part of California and even down to M 1.5 in some places, but the completeness magnitude increases close to the boundaries. In particular, there is a sharp decrease in the density of small events south of the Mexican border below 32° N. Correcting for catalog incompleteness is thus important when using small events to provide earthquake forecasts. In this study, we use the completeness map that was estimated by Werner *et al.* (2011) in each cell from the peak of a locally smoothed magnitude distribution. We also correct for changes in the detection threshold following large earthquakes, following the method described by Helmstetter *et al.* (2006) and

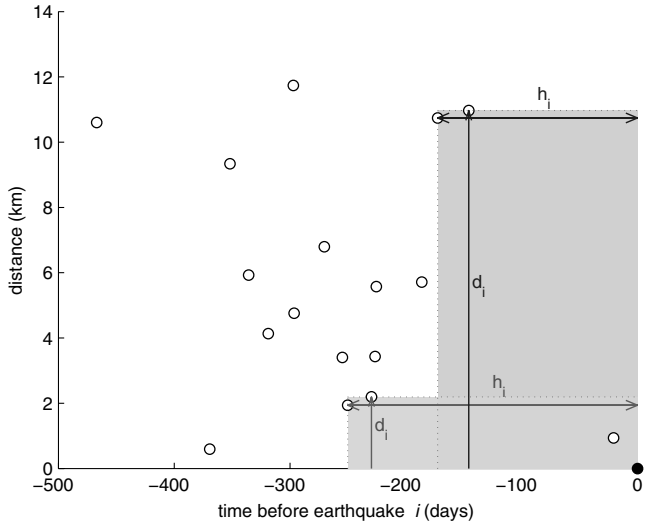

Figure 1. Illustration of the coupled near-neighbor method for estimating adaptive kernel bandwidths d_i and h_i in space and time, respectively, by minimizing the sum $h_i + ad_i$ subject to smoothing over k near-neighbors. Results are obtained for $k = 3$ and for different values of a: $a = 100$ (in light gray) and $a = 20$ day/km (in dark gray).

Werner *et al.* (2011). For each earthquake with $M > 5$, we use the relation given by Helmstetter *et al.* (2006) for the completeness magnitude M_c as a function of the magnitude M of the last large earthquake and the time T since this earthquake. We then assume that the magnitude distribution obeys the Gutenberg–Richter law in order to estimate the rate of $M \geq 2$ earthquakes from the rate of $M \geq M_c$ events.

To evaluate the predictive power of our models, we divide the catalog into two parts, a learning catalog and a target catalog. Only events in the learning catalog are used to compute the kernel bandwidths and the predicted earthquake density in all space–time bins. The target catalog is used to evaluate the model by computing the likelihood of the density using events in the target catalog. Following Werner *et al.* (2011), we use the time interval from 1 February 1981 until 31 December 2003 for the learning catalog. The target catalog contains all earthquakes from 1 January 2004 until 22 July 2011. Figure 2 illustrates the catalogs used in this study. The learning catalog with magnitudes $M \geq 2$ is shown in Figure 2a, and the target catalog with magnitudes $M \geq 4$ is shown in Figure 2b. We have also tested the models on a declustered target catalog to evaluate the influence of aftershocks on the predictive skill of the forecasts. For that purpose, we declustered using the method of Reasenberg (1985) with the same parameters and modifications that Helmstetter *et al.* (2007) used. The declustered catalog is shown in Figure 2c for the same target period (1 January 2004–22 July 2011) as for the (clustered) target catalog shown in Figure 2b.

Definition of the Grid

To adapt to the rules of CSEP's long-term earthquake predictability experiment in California, we produce forecasts on the grid covering the region defined by Schorlemmer and Gerstenberger (2007) for California shown in Figure 2 with a grid spacing of 0.1°. Because the cell size is much larger than the minimum value of the bandwidth (fixed to 0.5 km), we need to integrate the estimated seismicity rate in each cell to obtain the expected number of events per cell.

Correction for Catalog Incompleteness

Because of the variability of the detection threshold in time and space, we introduce a weight w in equation (2) associated with each earthquake (kernel), defined by

$$w_i = 10^{b(M_c(t_i,\mathbf{r_i})-M_d)}, \tag{6}$$

where b is the exponent of the Gutenberg–Richter magnitude distribution, $M_c(t_i, \mathbf{r_i})$ is the completeness magnitude estimated at the time and location of earthquake i, and M_d

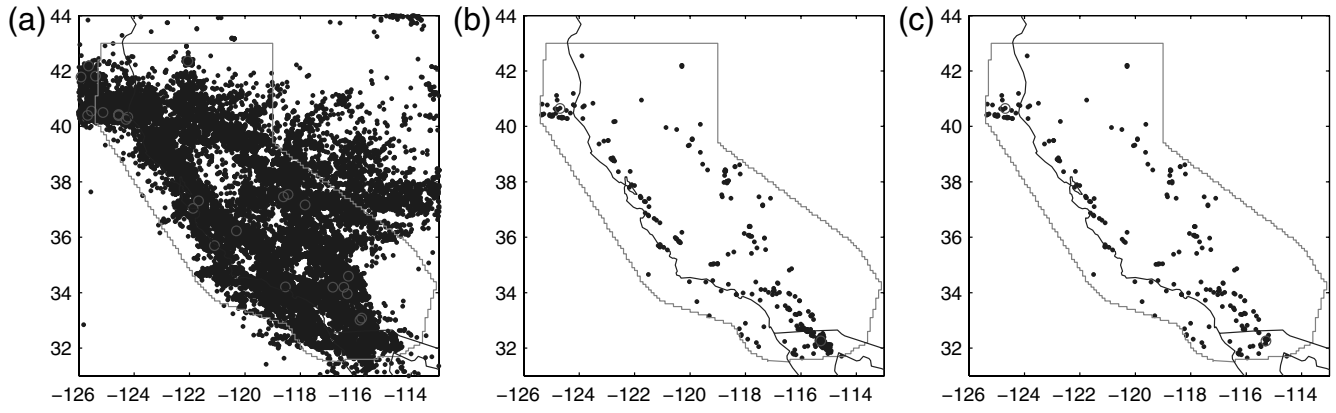

Figure 2. Map of earthquakes in the (a) learning, (b) target, and (c) declustered target catalogs. The gray line marks the limit of the CSEP testing area. The learning catalog includes 151,518 $M \geq 2$ earthquakes from 1 February 1981 until 31 December 2003. The target catalog and the declustered target catalog cover the time window from 1 January 2004 until 22 July 2011 and contain, respectively, $N = 423$ and $N = 228$ events with magnitude $M \geq 4$. Circles indicate $M \geq 6$ earthquakes.

is the minimum magnitude of earthquakes in the learning catalog. We use the completeness magnitudes calculated by Werner *et al.* (2011, their fig. 2b), which were calculated from the locally estimated magnitude distributions. In addition, M_c is modeled to increase after large $M \geq 5$ earthquakes using equation 15 of Helmstetter *et al.* (2006) to model increases in the detection threshold after mainshocks. Earthquakes with magnitudes smaller than the local completeness magnitude were deleted from the catalogs.

Magnitude Distribution

In this study, we do not attempt to estimate the future distribution of magnitudes; the likelihood function and probability gain only consider the spatial distribution of earthquakes. Nevertheless, we need a magnitude distribution in order to extrapolate the rate of $M \geq M_d$ events up to the magnitude threshold M_t of target earthquakes and to correct for catalog incompleteness using equation (6). For this goal, we use a simple Gutenberg–Richter magnitude distribution with $b = 1$. However, following Helmstetter *et al.* (2007) and Werner *et al.* (2011), we introduce a correction for the Geysers, a hydrogeothermal area in northern California. In this region, we assume a piecewise Gutenberg–Richter law with an exponent $b = 1$ for $M < 3.4$ and $b = 2$ for $M \geq 3.4$. The observed and modeled cumulative magnitude distributions are shown in Figure 3. When M_d is 2, the ratio of $M \geq M_t$ earthquakes for the piecewise linear magnitude distribution and the Gutenberg–Richter magnitude distribution with $b = 1$ is 0.13 for M_t of 4 and 0.013 for M_t of 5.

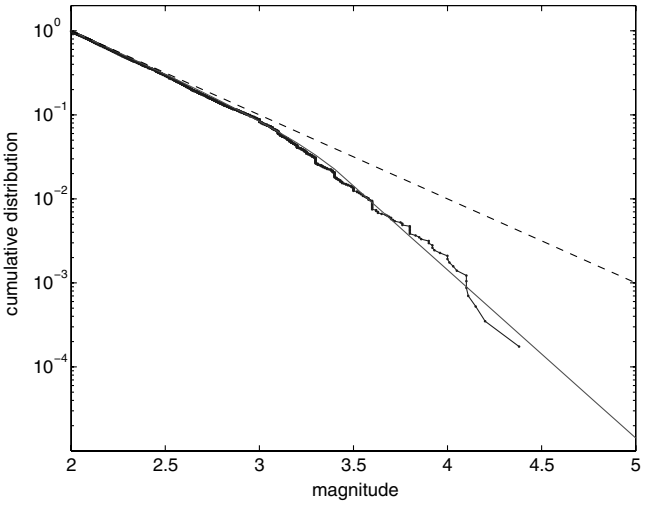

Figure 3. Observed (blue) and modeled (red line) cumulative magnitude distribution in the Geysers area (38.7 < latitude < 38.9°N and 122.7 < longitude < 122.9°W) using all earthquakes until 31 December 2003. The fit uses a piecewise Gutenberg–Richter law with an exponent $b = 1$ for $M < 3.4$ and $b = 2$ for $M \geq 3.4$.

Estimating Stationary Rates

We estimate the time series of the seismicity rate in each cell using adaptive kernels with a time step of 10 days. We also tried a longer time step of 100 days and obtained very similar results. The difference in probability gain was always smaller than one percent; the shorter time step usually produced better results. We also performed one simulation with an even shorter time step of one day and found that the gain increased by less than 0.04%. Therefore, we chose a step of 10 days as a good compromise between accuracy and computational effort. We did not try reducing this time step further because it considerably increases the computation time. The long-term rate $N_p(i_x, i_y)$ in each spatial cell (i_x, i_y) is then estimated by the median of the distribution of the 10-day rates in this cell. By choosing the median rate per cell, we effectively remove the influence of aftershock sequences without recourse to declustering algorithms. An illustration is shown in Figure 4 for a cell located at a latitude of 33.95° and a longitude of 116.45° W, close to the 1992 M 7.3 Landers earthquake.

Testing and Optimizing the Forecasts

As mentioned previously, here we only test the spatial distribution of seismicity and not the total number of predicted events or the magnitude distribution. We therefore normalize the predicted density so that the total expected number of events equals the total number of events in the target catalog. CSEP experiment rules for California fixed the minimum magnitude of target earthquakes to M_t of 4.95 for long-term forecasts. However, in this work we will also consider other values of M_t to test if the probability gain remains constant, which might suggest that the spatial distribution of large earthquakes is similar to that of smaller events.

The log-likelihood is defined by

$$L = \sum_{i_x=1}^{N_x} \sum_{i_y=1}^{N_y} \log p[N_p(i_x, i_y), n(i_x, i_y)], \qquad (7)$$

where $n(i_x, i_y)$ is the number of events that occurred in the cell (i_x, i_y). Assuming a Poisson process, the probability $p(N_p, n)$ of observing n events in one cell with a predicted rate N_p is given by

$$p(N_p, n) = \frac{N_p^n \exp(-N_p)}{n!}. \qquad (8)$$

Werner and Sornette (2008) and Werner *et al.* (2011) discussed the (in-)adequacy of this Poisson distribution for the variability in the number of earthquakes over the entire region and in each cell; we do not attempt to resolve these issues here.

The model parameters are optimized by maximizing the likelihood L using a simplex algorithm. The inverted parameters are the minimum density $R_{\min}$ in equation (2) and the parameters used to estimate the kernel bandwidths. If the bandwidths are estimated from a pilot estimate, the model

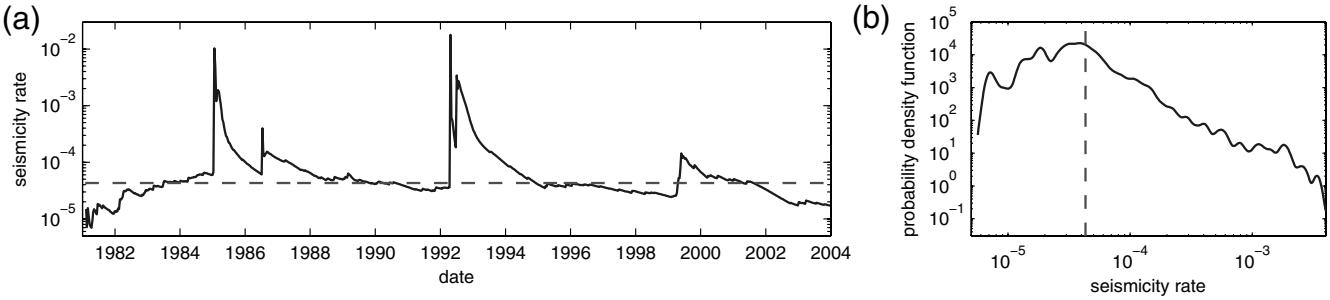

Figure 4. (a) Seismicity rate per day of $M \geq 4$ earthquakes as a function of time in the cell located at a latitude of 33.95° and a longitude of 116.45° W. The straight line represents the median value. Bandwidths were estimated using the coupled near-neighbor method with parameters $k = 14$ and $a = 226$ (model 3 in Table 1). The peak of the seismicity rate in 1985 corresponds to a small local cluster of $M < 4$ earthquakes. The seismicity rate also increases in 1992 after the M 6.2 Joshua Tree and M 7.3 Landers mainshocks and in 1999 after the M 7.1 Hector Mine earthquake. (b) The distribution of the seismicity rate is shown with the median value (dashed line).

parameters are the global values h_0 and d_0. For the coupled near-neighbor method, we invert for the number k of neighbors and the parameter a.

We define the gain G as the probability gain per target earthquake relative to a spatially uniform Poisson model (Kagan and Knopoff, 1977)

$$G = \exp\left(\frac{L - L_u}{N_t}\right). \tag{9}$$

The likelihood of the uniform model L_u is given by equations (7) and (8)

$$L_u = -N_t + \sum_{i_x=1}^{N_x} \sum_{i_y=1}^{N_y} n(i_x, i_y) \log(N_u) - \log[n(i_x, i_y)!]. \tag{10}$$

The rate in each cell is given by $N_u = N_t/N_c$, with N_c being the total number of cells in the testing area and N_t the total number of targets (we neglect differences in the surface areas between cells at different latitudes).

When comparing two models that have the same total number N_t, equation (9) can be simplified. Using equation (10) we obtain

$$\begin{aligned} L - L_u &= \sum_{i_x=1}^{N_x} \sum_{i_y=1}^{N_y} n(i_x, i_y) \log[N_p(i_x, i_y)/N_u] \\ &= \sum_{i=1}^{N_t} \log[N_p(i)/N_u], \end{aligned} \tag{11}$$

where $N_p(i)$ is the predicted rate in the bin in which earthquake i occurred. Here, the gain estimated from (9) represents the geometrical average of the ratio of the predicted rate for each model

$$G = \exp\left[\sum_{i=1,N_t} \frac{\log(N_p(i)/N_u)}{N_t}\right] = < N_p(i)/N_u >_{geo} . \tag{12}$$

When the total number of predicted events is conditioned on being equal to the observed number, maximizing the model likelihood is thus equivalent to maximizing the geometric average of the expected rates in cells with earthquakes.

Model Comparison using the *t*-test and *W*-test

We used the *t*-test and the *W*-test proposed by Rhoades *et al.* (2011) in order to compare different models. The *t*-test evaluates whether the information gain of a model is significantly different from that of another model. We have renormalized the predicted rate for all models so that the total expected number of events is equal to the observed number. We denote by $N_A(i)$ the expected rate at the location of earthquake i under model A and by $N_B(i)$ the rate under model B. In this context, the information gain per earthquake defined by equation 17 of Rhoades *et al.* (2011) reduces to

$$I = \frac{1}{N_t} \sum_{i=1,N_t} x_i, \tag{13}$$

where $x_i = \log[N_A(i)/N_B(i)]$. The information gain I is related to the probability gain defined by (9) of models A and B by $I = \log(G_A/G_B)$.

The sample variance of x_i is given by equation 18 of Rhoades *et al.* (2011)

$$s^2 = \frac{1}{N_t - 1} \sum_{i=1,N_t} x_i^2 - \frac{1}{N_t^2 - N_t} \left(\sum_{i=1,N_t} x_i\right)^2. \tag{14}$$

The so-called T-value is then defined by Rhoades *et al.* (2011)

$$T = \frac{I\sqrt{N_t}}{s}. \tag{15}$$

If x_i are independent and obey a normal distribution, then T has a student distribution with $N_t - 1$ degrees of freedom. For large N_t, the student distribution rapidly converges to a normal distribution. The information gain of model A can then be considered significantly greater than for model B if $T > 2$ at the 95% confidence interval.

The Wilcoxon signed-rank test or *W*-test (e.g., Siegel, 1956) can also be used in the case where x_i are not normally

distributed but remain symmetric and independent (Rhoades *et al.*, 2011). This test evaluates whether the median value of x_i is significantly different from 0; that is, if the predicted rates $N_A(i)$ differ significantly from $N_B(i)$. This test returns a probability p_W of observing a value larger than the median of x_i by chance. A value $p_W < 0.05$ indicates that the median value of x_i is significantly different from 0.

Results

Comparison of the Iterative and Near-Neighbor Methods

We evaluated the two methods for estimating kernel bandwidths that we mentioned previously and varied the time and magnitude intervals for the learning and target catalogs. We varied the minimum magnitude M_d for events in the learning catalog and the minimum magnitude M_t for events in the target catalog. The results are summarized in Table 1.

The results shown in Table 1 are slightly better with the near-neighbor method for small magnitudes M_d of the learning catalog, while the pilot density performs generally better for $M_d \geq 3$. Our final, preferred forecast for future seismicity corresponds to model 4 in Table 1. It has the same parameters as model 3 (which boasts the highest gain for $M_t \geq 4$), but uses all available data for the learning catalog, from 1 February 1981 until 22 July 2011. Because the learning and target catalogs partly overlap, the prediction gain $G = 5.27$ is larger than the value $G = 4.60$ obtained for model 3.

Figure 5 compares the kernel bandwidths in time and space estimated using the two methods. The coupled near-neighbor method provides on average smaller values of the temporal bandwidth h and the spatial bandwidth d. In the pilot-density method, h is proportional to d, except for values of $d < 0.5$ km that were replaced by 0.5 km to account for location accuracy. In contrast, when using the near-neighbor method, there is only a weak positive correlation between h and d (Fig. 5c). Letting the data speak for themselves thus seems to suggest that forcing a proportionality between temporal and spatial smoothing may not be adequate for representing seismicity fluctuations.

Comparison with the Model of Werner *et al.* (2011)

We compare the results with a long-term forecast generated with the method described by Werner *et al.* (2011), hereafter referred to as W11. This model uses the method of Reasenberg (1985) for declustering the (learning) seismicity and adaptive kernels in space to smooth the location of past earthquakes.

To benchmark the predictive skill of the new methods, we regenerated a forecast using the method of Werner *et al.*

Table 1
Definition of the Catalogs, Model Parameters, and Results*

	Learning Catalog		Target Catalog		Parameters					Results
Model	M_d	N_l	M_t	N_t	k	a	h_0	d_0	$N_{\min}$	G
1	2.0	151518	2.0	26145	5	382	–	–	1.2×10^{-3}	5.50
2	2.0	151518	3.0	3875	13	247	–	–	4.6×10^{-3}	4.41
3	2.0	151518	4.0	423	14	226	–	–	8.2×10^{-5}	4.60
4	2.0	188005	4.0	423	14	226	–	–	8.2×10^{-5}	5.27
5	2.0	151518	5.0	36	7	365	–	–	6.6×10^{-5}	7.40
6	3.0	16502	3.0	3875	7	151	–	–	1.4×10^{-4}	3.98
7	3.0	16502	4.0	423	2	124	–	–	7.1×10^{-6}	4.37
8	3.0	16502	5.0	36	1	547	–	–	1.0×10^{-5}	8.69
9	4.0	1700	4.0	423	1	177	–	–	1.3×10^{-5}	3.31
10	4.0	1700	5.0	36	1	292	–	–	1.3×10^{-7}	5.85
11	2.0	151518	2.0	26145	–	–	62.1	0.27	1.7×10^{-3}	5.32
12	2.0	151518	3.0	3875	–	–	59.9	0.78	1.8×10^{-5}	4.28
13	2.0	151518	4.0	423	–	–	45.1	0.25	3.1×10^{-6}	4.38
14	2.0	151518	5.0	36	–	–	16.9	0.17	1.6×10^{-6}	7.29
15	3.0	16502	3.0	3875	–	–	33.3	0.52	8.4×10^{-4}	4.08
16	3.0	16502	4.0	423	–	–	92.4	0.48	6.5×10^{-5}	4.41
17	3.0	16502	5.0	36	–	–	135	0.53	3.6×10^{-6}	9.29
18	4.0	1700	4.0	423	–	–	149	2.14	6.3×10^{-5}	3.60
19	4.0	1700	5.0	36	–	–	177	1.24	8.8×10^{-5}	7.06

*N_l and N_t are the total number of events in the learning and target catalogs, respectively. k, a, h_0 and d_0 are smoothing parameters. $N_{\min}$ is a minimum constant rate, in units of number of events with $M \geq M_t$ per day in the testing area. G is the probability gain per earthquake relative to a Poisson model with a uniform rate. Time variables are given in units of days and distances in km. The first 10 models were constructed with the coupled near-neighbor method, while the subsequent models are based on the pilot density. Model 4 is our preferred model for future seismicity, which uses all available data for the learning catalog from 1 February 1981 until 22 July 2011. All other models use only data until 31 December 2003 for the learning catalog, so that there is no overlap with the target catalog.

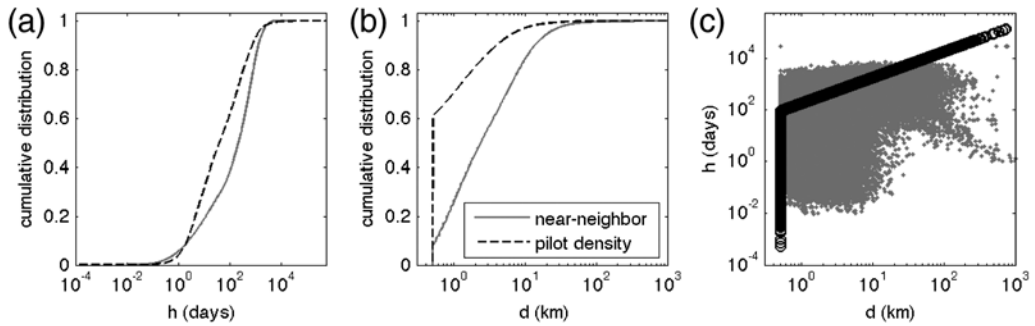

Figure 5. (a) Comparison between the bandwidths in time h and (b) space d estimated using the coupled near-neighbor method (model 3 in Table 1, solid line) and the method based on a pilot density (model 13 in Table 1, dashed line). (c) Relation between the temporal and spatial bandwidths for each earthquake, using the coupled near-neighbor method (points) and the pilot density (black circles). Values of d smaller than 0.5 km were replaced by 0.5 km to account for location accuracy.

(2011) using a (declustered) input catalog that covers the same learning period (up to 31 December 2003) as the input for the two new methods. We have optimized the smoothing parameter k used in that study, defined as the distance to the kth nearest neighbor, using the same target catalog as in our new models. Werner *et al.* (2011) preferred $k = 10$ based on the then available data, but optimizing the forecast for the present target window yields $k = 26$, yielding a smoother forecast. Figures 6, 7, and 8 compare new forecasts with the long-term forecast W11. All models were built using data

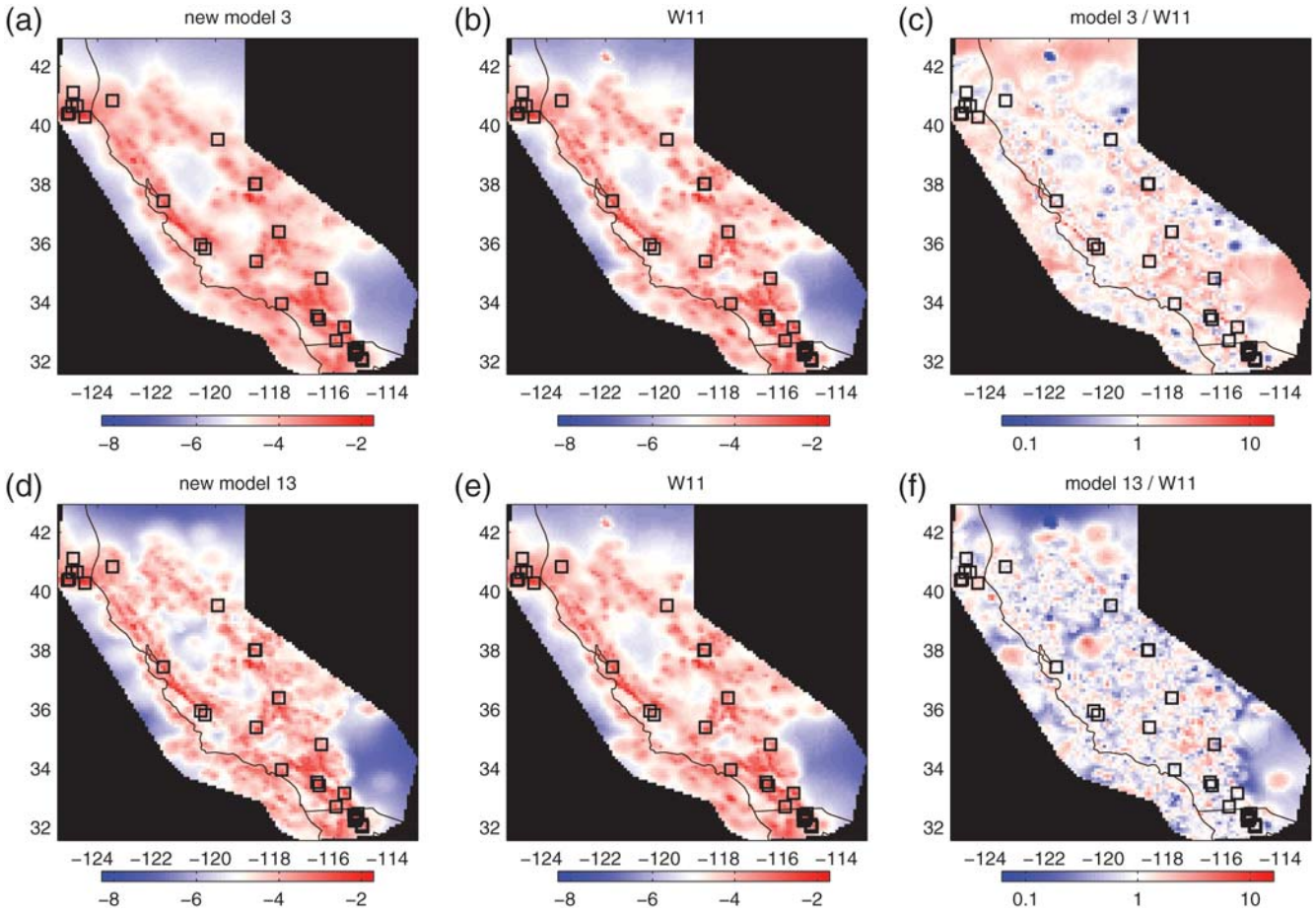

Figure 6. Comparison of our models (a) 3 and (d) 13 with (b and e) model W11. All models use the same input catalog from 1981 to 2004 above $M \geq 2$ to estimate the long-term rate. The color-bar gives the predicted number of events per cell and per day above $M_t = 4$ in log scale. Black squares indicate $M \geq 5$ earthquakes that occurred from 1 January 2004 until 22 July 2011. Ratio of predicted rates according to our model (c) 3 or (e) 13 and model W11, in log scale.

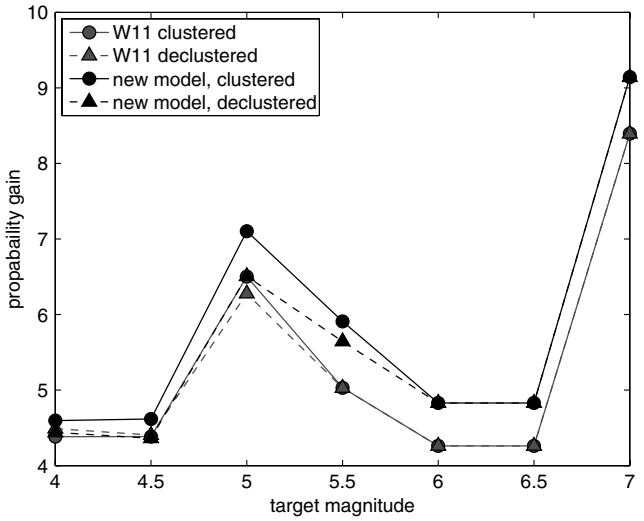

Figure 7. Comparison between long-term model 3 from Table 1 and model W11, tested on both the original and a declustered version of target catalog from 1 January 2004 to 22 July 2011, for different values of the minimum magnitude M_t.

from 1981 until 2004 with magnitudes $M \geq 2$ to calculate the long-term rate of $M \geq 4$ earthquakes. On average, model 3 based on the near-neighbor method is smoother than W11. In comparison, model 13 based on the pilot estimate has more contrast; it is more focused in zones of dense seismicity and predicts a lower rate close to the borders of the testing area. The zone where the new models differ most from model W11 is close to the point 42.4° N and 122.1° E, where

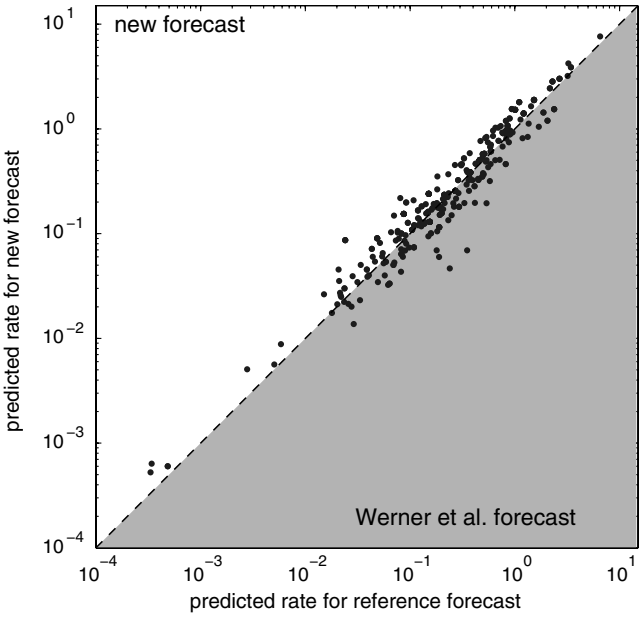

Figure 8. Comparison between the rate $N_p(i)$ predicted by our model in the bins where each target occurred (vertical axis) as a function of the rate predicted by model W11 (horizontal axis). The target catalog includes 423 $M \geq 4$ earthquakes from 1 January 2004 until 22 July 2011 without declustering.

the rate predicted by W11 is more than 1000 times greater. An M 6 mainshock occurred in this area in 1993, followed by numerous aftershocks, not all of which seem to have been removed by the declustering method of Reasenberg (1985) used to generate W11. The seismicity rate in this zone in 2011 remains higher than it was before the mainshock. However, the period of the learning catalog preceding the mainshock is longer than the one after it; therefore, the median value of the rate in this area computed from the learning catalog is close to the rate observed before the mainshock, which is much lower than the rate estimated by W11. Similar reasons hold for other regions of large differences. In hindsight, given the present, slightly better results of the new methods, one might conclude that the Reasenberg declustering algorithm failed to remove sufficient aftershocks and is not optimized for generating forecasts.

Figure 7 compares the probability gains G of models 3 and W11. For both models, G ranges between 4 and 8 as the minimum magnitude of targets M_t varies between 4 and 7 (and the spatial forecast is held fixed). The gain also shows no clear trend with M_t, suggesting that moderate and large earthquakes have the same spatial distribution. We found slightly larger values of G for our new model compared with W11.

The target catalog includes a large cluster of earthquakes triggered by the M 7.2 Baja, California, earthquake of 10 April 2010, raising the question of the influence of aftershocks on the results. To address this, we also tested a declustered target catalog shown in Figure 2c and found small variations of G. Model W11 better predicts the locations of (presumed) background earthquakes than our new model when considering declustered target earthquakes with $M \geq 4$ or $M \geq 4.5$, but our new model is still better for M_t of 5 or greater.

In Figure 8, we show that the rate $N_p(i)$ predicted by model 3 for each quake is larger than the rate $N_W(i)$ predicted by W11 for 250 out of 423 earthquakes. The arithmetic and geometric average of N_p/N_W are both larger than one, hence the larger gain G for our new model. Nonetheless, the differences between the two models in terms of the expected rates in cells with earthquakes is usually small. The new model is slightly better for both very large and very small values of the predicted rate. One earthquake is even five times more likely according to model W11 relative to the new model. For intermediate values ($0.01 < N_W < 1$), the predicted rates by both models are on average very close.

t-test and W-test

We used the t- and W-tests to compare the predictive skill of our new models with W11. We tested variants of the new methods (pilot estimate and nearest neighbor) at different target magnitudes and on both the original and a declustered target catalog. In Table 2, we summarize those results. Model A (first column) can be considered as significantly better than model B (second column) at the 95%

confidence interval if $T > 2$ and $p_W < 0.05$. For large target earthquakes $M_t \geq 5$, the number of events is too small to reject any forecast (i.e., the differences between the forecasts is too small). For smaller thresholds, our new model 3 is significantly better than W11 when M_t equals 3 or 4. But W11 is better than the new model in forecasting the declustered target catalog, although the difference is not significant. In contrast, the pilot-estimate method yields better results than the near-neighbor method for the declustered catalog. The gain of model 13 is significantly larger than the gain of model 3 when M_t equals 3 or 4, and larger than the gain of W11 when M_t is 3.

Concentration Plot

Figure 9 is a concentration plot (Rong and Jackson, 2002) that compares the distribution of the predicted events with observed number of events and allows for an evaluation of the earthquake forecasts with respect to the observations. The forecast rates of all spatial bins are sorted in ascending order, and their cumulative distribution is then plotted. To compare the forecast to observations, the number of observed earthquakes is sorted according to the same order as the forecast rates and their cumulative distribution plotted. A concentration diagram provides a visual inspection of the concentration (or smoothness) of a forecast, and whether this is matched by the observed distribution of earthquakes. A good model concentrates most earthquakes in cells with high predicted rate toward the right part of the diagram. Moreover, if the model was fully consistent with the observations, the distribution of the observed number of events (red line) would match the predicted number of events (blue line).

However, we see a significant difference between these two curves. Our model predicts too few events in cells where the predicted rate N_p is smaller than 0.05, while it overpredicts the number of events in cells with a predicted rate in the range between 0.05 and 1. This could suggest that the model is not smooth enough, even though the smoothing parameters were optimized to maximize the forecast likelihood. To test if the difference is significant, we generated 100 synthetic catalogs consistent with our model (light blue lines in Fig. 9). However, as is apparent from the difference between the red line (observations) and the blue lines (synthetics), the random fluctuations observed in the synthetics cannot explain the deviation between the model and the observations.

This difference can be partly explained by aftershocks: when using the declustered target catalog (green line in Fig. 9), the difference between the model and the observations decreases significantly. But the green curve is still a little outside of the confidence interval given by the synthetic catalogs in areas of very dense seismicity (predicted rate larger than 1). This suggests that the model is a little too smooth in very active zones. This inconsistency might also suggest that the background seismicity is not stationary, but that there are slow variations of the tectonic loading. We repeated this analysis with model W11 and obtained very similar results.

Table 2
Results of the *t*-test and *W*-test for Comparing Different Models

Model *A*	Model *B*	M_t	N_t	Declustered	G_A	G_B	T	p_W
n.n.*	W11	3.0	3875	no	4.36	4.13	8.20	2×10^{-20}
n.n.*	W11	4.0	423	no	4.60	4.38	2.73	7×10^{-4}
n.n.	W11	5.0	36	no	7.10	6.50	1.75	0.01
n.n.	W11	3.0	2002	yes	4.65	4.70	−1.21	0.28
n.n.	W11	4.0	228	yes	4.44	4.49	−0.45	0.68
n.n.	W11	5.0	23	yes	6.51	6.28	0.53	0.14
p.l.	W11	3.0	3875	no	4.11	4.13	−1.12	0.79
p.l.	W11	4.0	423	no	4.38	4.38	−0.05	0.72
p.l.	W11	5.0	36	no	6.95	6.50	1.09	0.11
p.l.*	W11	3.0	2002	yes	4.94	4.70	4.87	2×10^{-15}
p.l.	W11	4.0	228	yes	4.73	4.49	1.83	4×10^{-3}
p.l.	W11	5.0	23	yes	6.71	6.28	0.75	0.25
n.n.*	p.l.	3.0	3875	no	4.36	4.11	8.99	7×10^{-9}
n.n.	p.l.	4.0	423	no	4.60	4.38	2.61	0.12
n.n.	p.l.	5.0	36	no	7.10	6.95	0.44	0.94
n.n.	p.l.*	3.0	2002	yes	4.65	4.94	−6.93	6×10^{-22}
n.n.	p.l.*	4.0	228	yes	4.44	4.73	−2.41	2×10^{-4}
n.n.	p.l.	5.0	23	yes	6.51	6.71	−0.50	0.33

Results of the t and W tests for comparing a model A with another model B for different values of the minimum target magnitude M_t, with or without declustering the target catalog. In columns 1 or 2, n.n. indicates the near-neighbor method with the parameters of model 3, p.l. refers to the pilot-estimate method with the parameters of model 13 in Table 1, and W11 refers to the long-term model of Werner *et al.* (2011). N_t is the number of target earthquakes. T and p_W are test results described in the *t*-test and *W*-test section. The probability gain G_A for model A is significantly larger than G_B if $T > 2$ and $p_W < 0.05$; in this case, asterisks indicate the best model. The target catalog contains all $M \geq M_t$ earthquakes from 1 January 2004 until 22 July 2011 with or without declustering.

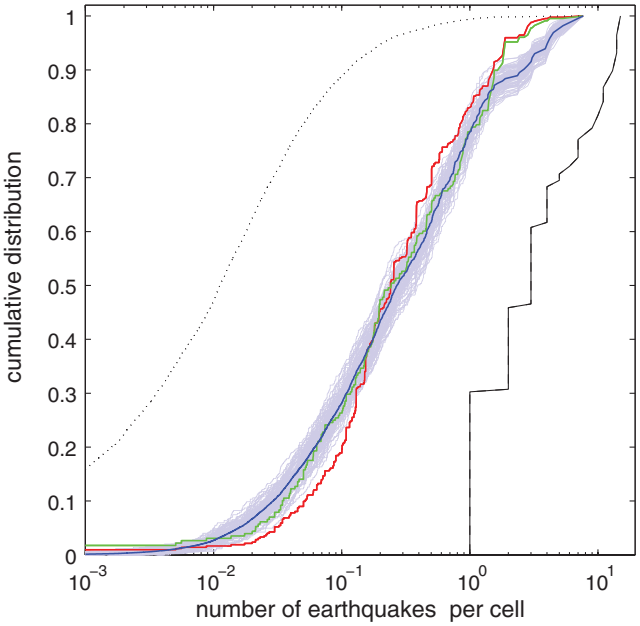

Figure 9. Cumulative distribution of the number of events (observed or forecasted) as a function of the minimum value of the forecasted rate calculated using our model 3. The blue curve represents the cumulative distribution of the expected number of events per cell. The red and green curves show respectively the cumulative observed number of events per cell in the target catalog and in the declustered catalog as a function of the predicted rate. These curves should overlap if the model were fully consistent with the data. Light blue curves show the observed number of events in 100 synthetic catalogs, generated using a Poisson process with an expected rate in each cell given by our model $N_p(i_x, i_y)$. The dotted line shows the expected number of events assuming a uniform Poisson process. In addition, the solid black line shows the cumulative number of observed events as a function of the observed number of events in each cell, this represents the curve that would be obtained for a perfect model with $N_p(i_x, i_y) = n(i_x, i_y)$.

Our model may still be improved by including a spatially variable b-value, in order to account for the spatial variability of the magnitude distribution (e.g., Wiemer and Schorlemmer, 2007). We could also use a different kernel function in time and space. For instance, a power-law kernel with heavier tails would result in a smoother density in cells with intermediate and very little activity, and more focused density in zones of very dense seismicity.

Conclusion

We presented two new methods for long-term earthquake forecasting based on adaptive kernels, which are used to smooth past seismicity in both space and time. The long-term seismicity rate in each cell is estimated from the median value of the seismicity rate in that cell, thereby avoiding the need for a controversial declustering algorithm. Most other long-term smoothed seismicity forecasts, including ours (Helmstetter *et al.*, 2007; Werner *et al.*, 2011), are based on spatially smoothing declustered catalogs using uniform or adaptive kernels.

One of our previous long-term forecasts was submitted to the prospective, five-year Regional Earthquake Likelihood Models (RELM) experiment in California, which started in January 2006 and ended in 2011 (Field, 2007). Schorlemmer *et al.* (2010) evaluated the models after the first half of the testing period and found that the model of Helmstetter *et al.* (2007) performed better than its competitors. The final evaluation confirmed this result (Zechar *et al.*, unpublished manuscript, 2012). The retrospective testing performed here suggests that our new method might be even better: When evaluated using the catalog of $M \geq 4$ earthquakes for the period 2004–2011, our new model performs slightly better than model W11 (itself an improvement of the model by Helmstetter *et al.*, 2007) when both models are built from the same learning catalog (1981–2004 and $M \geq 2$). When compared with the model of Helmstetter *et al.* (2007) using target earthquakes with $M \geq 4$ or $M \geq 5$ for the period 1 January 2006–15 July 2011, the new model produces a slightly larger gain. The new forecast is a little smoother and seems to better remove the influence of previous large aftershock sequences.

While our new model slightly improves the forecast likelihood, it is not fully consistent with the data, as can be shown from the concentration plot. The difference between the observations and the model is greatly reduced but not totally suppressed by using a declustered catalog. This suggests that the model may be improved by better optimizing the smoothing or by including spatial variations of the magnitude distribution. In particular, it would be interesting to test different kernels in space and time, for instance, replacing the Gaussian kernel by a power law. To demonstrate the (truly prospective) predictive value of the new forecasting method presented here, we have submitted this new long-term forecast to CSEP experiments in California. The prospective test will start on 1 January 2013. We will also apply our model to other areas around the globe and submit those models to CSEP.

Data and Resources

We used the Advanced National Seismic System (ANSS) earthquake catalog made publicly available by the Northern California Earthquake Data Center at www.ncedc.org (last accessed July 2011) in the period from 1 February 1981 until 22 July 2011 with magnitude $M \geq 2$ and in the spatial region defined by the RELM collection region, defined in table 2 by Schorlemmer and Gerstenberger (2007).

Acknowledgments

The authors thank Jeremy Zechar and an anonymous reviewer for carefully reading the paper and providing many constructive suggestions. A.H. is supported by the French National Research Agency under Grant ASEISMIC. M.J.W. was partially supported by the Southern California Earthquake Center (SCEC). SCEC is funded by NSF Cooperative Agreement EAR-0106924 and USGS Cooperative Agreement 02HQAG0008. This research was supported by the Southern California Earthquake Center. SCEC is funded by NSF Cooperative Agreement EAR-0529922 and USGS Cooperative Agreement 07HQAG0008. The SCEC contribution number for this paper is 1667.

References

Abramson, I. S. (1982). On bandwidth variation in kernel estimates: A square root law, *Ann. Stat.* **10,** 1367–1223.

Adelfio, G., and M. Chiodi (2010). Diagnostics for nonparametric estimation in space-time seismic processes, *J. Environ. Stat.* **1,** no. 2, 1–13.

Choi, E., and P. Hall (1999). Nonparametric approach to the analysis of spacetime data on earthquake occurrences, *J. Comput. Graph. Stat.* **8,** 733–748.

Field, E. H. (2007). Overview of the Working Group for the Development of Regional Earthquake Likelihood Models (RELM), *Seism. Res. Lett.* **78,** no. 1, 7–16.

Frankel, A., C. Mueller, T. Barnhard, D. Perkins, E. V. Leyendecker, N. Dickman, S. Hanson, and M. Hopper (1997). Seismic-hazard maps for California, Nevada, and Western Arizona/Utah, *U.S. Geol. Surv. Open-File Rept. 97-130*.

Gardner, J. K., and L. Knopoff (1974). Is the sequence of earthquakes in Southern California, with aftershocks removed, Poissonian? *Bull. Seismol. Soc. Am.* **64,** 363–367.

Hainzl, S., F. Scherbaum, and C. Beauval (2006). Estimating background activity based on interevent-time distribution, *Bull. Seismol. Soc. Am.* **96,** no. 1, 313–320, doi: 10.1785/0120050053.

Helmstetter, A., Y. Y. Kagan, and D. D. Jackson (2006). Comparison of short-term and time-independent earthquake forecast models for southern California, *Bull. Seismol. Soc. Am.* **96,** doi: 10.1785/0120050067.

Helmstetter, A., Y. Y. Kagan, and D. D. Jackson (2007). High-resolution time-independent grid-based forecast for $M \geq 5$ earthquakes in California, *Seismol. Res. Lett.* **78,** 78–86, doi: 10.1785/gssrl.78.1.78.

Jordan, T. H. (2006). Earthquake predictability: Brick by brick, *Seismol. Res. Lett.* **77,** no. 1, 3–6.

Kagan, Y. Y. (1991). Likelihood analysis of earthquake catalogs, *Geophys. J. Int.* **106,** 135–148.

Kagan, Y., and L. Knopoff (1977). Earthquake risk prediction as a stochastic process, *Phys. Earth Planet. In.* **14,** 97–108.

Kagan, Y. Y., and D. D. Jackson (1994). Long-term probabilistic forecasting of earthquakes, *J. Geophys Res.* **99,** no. B7, 13,685–13,700, doi: 10.1029/94JB00500.

Marsan, D., and O. Lengliné (2008). Extending earthquakes' reach through cascading, *Science* **319,** no. 5866, 1076–1079, doi: 10.1126/science.1148783.

Reasenberg, P. A. (1985). Second-order moment of central California seismicity, *J. Geophys. Res.* **90,** no. B7, 5479–5496.

Rhoades, D. A., D. Schorlemmer, M. Gerstenberger, A. Christophersen, J. D. Zechar, and M. Imoto (2011). Efficient testing of earthquake forecasting models, *Acta Geophys.* **59,** no. 4, 728–747, doi: 10.2478/s11600-011-0013-5.

Rong, Y., and D. D. Jackson (2002). Earthquake potential in and around China: Estimated from past earthquakes, *Geophys. Res. Lett.* **29,** no. 16, 1780, doi: 10.1029/2002GL015297.

Schorlemmer, D., and M. Gerstenberger (2007). RELM testing center, special issue, *Seismol. Res. Lett.* **78,** no. 1, 30–36.

Schorlemmer, D., J. D. Zechar, M. J. Werner, E. Field, D. D. Jackson, and T. H. JordanThe RELM Working Group (2010). First results of the regional earthquake likelihood models experiment, *Pure Appl. Geophys.* **167,** no. 8–9, 859–876, doi: 10.1007/s00024-010-0081-5.

Siegel, S. (1956). *Nonparametric Statistics for the Behavioral Sciences*, McGraw-Hill, New York, 312 pp.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London, 175 pp.

Stock, C., and E. G. C. Smith (2002a). Adaptive kernel estimation and continuous probability representation of historical earthquake catalogs, *Bull. Seismol. Soc. Am.* **92,** no. 3, 904–912.

Stock, C., and E. G. C. Smith (2002b). Comparison of seismicity models generated by different kernel estimations, *Bull. Seismol. Soc. Am.* **92,** no. 3, 913–922.

Toda, S., R. S. Stein, G. C. Beroza, and D. Marsan (2012). Aftershocks halted by static stress shadows, *Nat. Geosci.* **5,** doi: 10.1038/ngeo1465.

Van Stiphout, T., D. Schorlemmer, and S. Wiemer (2011). The effect of uncertainties on estimates of background seismicity rate, *Bull. Seismol. Soc. Am.* **101,** no. 2, 482–494, doi: 10.1785/0120090143.

Werner, M. J., and D. Sornette (2008). Magnitude uncertainties impact seismic rate estimates, forecasts, and predictability experiments, *J. Geophys. Res.* **113,**, B08302, doi: 10.1029/2007JB005427.

Werner, M. J., A. Helmstetter, D. D. Jackson, and Y. Y. Kagan (2011). High-resolution long-term and short-term earthquake forecasts for California, *Bull. Seismol. Soc. Am.* **101,** no. 4, 1630–1648, doi: 10.1785/0120090340.

Werner, M. J., A. Helmstetter, D. D. Jackson, Y. Y. Kagan, and S. Wiemer (2010). Adaptively smoothed seismicity earthquake forecasts for Italy, *Ann. Geophys.* **53,** no. 3, 107–116, doi: 10.4401/ag-4839.

Werner, M. J., J. D. Zechar, W. Marzocchi, and S. Wiemer (2010). Retrospective evaluation of the five-year and ten-year CSEP-Italy earthquake forecasts, *Ann. Geophys.* **53,** no. 3, 11–30, doi: 10.4401/ag-4840.

Wiemer, S., and D. Schorlemmer (2007). ALM: An asperity-based likelihood model for California, *Seismol. Res. Lett.* **78,** 134–140.

Zechar, J. D., and T. H. Jordan (2010). Simple smoothed seismicity earthquake forecasts for Italy, *Ann. Geophys* **53,** no. 3, doi: 10.4401/ag-4845.

Zechar, J. D., D. Schorlemmer, M. Liukis, J. Yu, F. Euchner, P. J. Maechling, and T. H. Jordan (2010). The Collaboratory for the Study of Earthquake Predictability perspective on computational earthquake science, *Concurrency Comput. Pract. Ex.* **22,** 1836–1847, doi: 10.1002/cpe.1519.

Zhuang, J., Y. Ogata, and D. Vere-Jones (2002). Stochastic declustering of spacetime earthquake occurrences, *J. Am. Stat. Assoc.* **97,** 369–380.

Zhuang, J., Y. Ogata, and D. Vere-Jones (2004). Analyzing earthquake clustering features by using stochastic reconstruction, *J. Geophys. Res.* **109,** B05301, doi: 10.1029/2003JB002879.

ISTerre
Université de Grenoble 1
CNRS, BP 53
F-38041 Grenoble
France
ahelmste@ujf-grenoble.fr
(A.H.)

Department of Geosciences
Princeton University
Princeton, New Jersey 08544
(M.J.W.)

Manuscript received 20 February 2012